DOCUMENT RESUME

ED 026 078 LI 000 979

By-Agenbroad, James E; And Others

Systems Design and Pilot Operations of a Regional Center for Technical Processing for the Libraries of the New England State Universities. NELINET, New England Library Information Network. Progress Report, July 1, 1967 - March 30, 1968. Volume I.

Inforonics, Inc., Cambridge, Mass.
Spons Agency-New England Board of Higher Education, Wellesley, Mass.
Pub Date 5 Apr 68
Contract-CLR-385
Note-41p.; Vol. II, Appendices, is LI 000 980
EDRS Price MF-$0.25 HC-$2.15
Descriptors-Automation, Cataloging, *Centralization, Information Processing, Library Acquisition, *Library Networks, *Library Technical Processes, Pilot Projects, Regional Programs, *Systems Development, *University Libraries
Identifiers-NELINET, *New England Library Information Network

The New England Library Information Network (NELINET) entered the pilot project stage late in 1967 when the University of New Hampshire became the first member of the network to be connected to the processing center in Cambridge, Massachusetts. The five other New England state university libraries were ready to be connected at the time of this report. The report focuses on (1) the systems analysis and program planning for the project, and (2) the services available under the pilot operation. Section One is an introduction and summary. Section Two covers initial work with catalog date file searching. Section Three deals with acquisitions processing. Section Four is a demonstration of services carried out in connection with the pilot operation. Volume II (LI 000 980) contains appendices giving much of the technical detail of project activities. (CC)

LI 000979

ERIC/CLIS
SEP 24'68
UNIVERSITY OF MINNESOTA

ED026078

Inforonics*

NELINET - NEW ENGLAND LIBRARY INFORMATION NETWORK

PROGRESS REPORT

JULY 1, 1967 - MARCH 30, 1968

SYSTEMS DESIGN AND PILOT OPERATION OF A REGIONAL CENTER FOR TECHNICAL PROCESSING FOR THE LIBRARIES OF THE NEW ENGLAND STATE UNIVERSITIES

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

VOLUME I

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

PREPARED BY

JAMES E. AGENBROAD, LAWRENCE F. BUCKLAND, ANN T. CURRAN,
DONALD D. HODGINS, WILLIAM R. NUGENT, ROBERT H. SIMMONS

SUBMITTED TO

THE NEW ENGLAND BOARD OF HIGHER EDUCATION

FINAL REPORT

CONTRACT NO. CLR - 385

APRIL 5, 1968

LI 000979

*new concepts in information organization, processing, and presentation

INFORONICS, INC.
806 MASSACHUSETTS AVENUE, CAMBRIDGE, MASSACHUSETTS 02139, TEL. (617) 547-1750
146 MAIN STREET, MAYNARD, MASSACHUSETTS 01754, TEL. (617) 897-8815
927 15TH STREET, N. W., WASHINGTON, D. C. 20005, TEL. (202) 638-6862

TABLE OF CONTENTS

		PAGE
1.	INTRODUCTION	1
2.	CATALOG DATA FILE SEARCHING: PLAN FOR EXPERIMENTS AND SYSTEM DEVELOPMENT	3
	2.1 INTRODUCTION	3
	2.2 DEVELOPMENT PLAN FOR FILE SEARCHING	3
	2.3 GENERAL SEARCHING METHOD	4
	2.4 TITLE SEARCHING	6
	2.4.1 TITLE KEY DETERMINATION	6
	2.4.1.1 TITLE WORD SELECTION	7
	2.4.2 KEY CODING	7
	2.4.3 TITLE DIRECTORY-PRELIMINARY	14
	2.5 AUTHOR SEARCHING	14
	2.5.1 TRANSITION DISTANCE CODING AND ALPHACHECK	17
	2.5.2 AUTHOR DIRECTORY-PRELIMINARY	19
	2.6 CONCLUSION	19
3.	TASK 3: ACQUISITION PROCESSING	23
	3.1 ANALYSIS OF EXISTING ACQUISITIONS SYSTEMS	23
	3.1.1 FORMS	23
	3.1.2 FILES	30
	3.1.3 TIE-IN WITH THE CATALOGING PROCESS	30
	3.2 DESIGN OF THE MECHANIZED ACQUISITIONS SYSTEM	32
4.	TASK 4: DEMONSTRATION OF SERVICES	34

TABLE OF CONTENTS

APPENDIX A - ACQUISITIONS FLOW CHARTS

APPENDIX B - CURRENT OPERATIONS QUESTIONNAIRE

APPENDIX C - AN ALGORITHM FOR SPLITTING THE LC CALL NUMBER

APPENDIX D - ANALYSIS OF THE MARC II FORMAT

APPENDIX E - PRODUCTION PROBLEMS AND DECISIONS

APPENDIX F - OPERATING PROCEDURES FOR INFORMATION TRANSMITTAL IN THE NEW ENGLAND LIBRARY INFORMATION NETWORK

APPENDIX G - COMPRESSION WORD CODING TECHNIQUES: TRANSITION DISTANCE CODING, ALPHACHECK, RECURSIVE DECOMPOSITION AND SOUNDEX

APPENDIX H - SAMPLE CARDS AND LABELS

I. INTRODUCTION AND SUMMARY

NELINET progressed from a development project to one of pilot operation in the last days of 1967, when the University of New Hampshire became the first member of the network to be connected to the processing center in Cambridge, Massachusetts. In the days between then and this writing, extensive shakedown and tests of the cataloging services have been performed, so that we are now ready to connect the remaining five libraries of the New England State Universities. This is now being scheduled.

Before, and continuing after the initiation of pilot operation, extensive development work has been in progress on additional aspects of the network and its services, that in total will provide complete technical processing services to member libraries in New England. This is a technical report in depth on the development work to date, and on the pilot operation.

The background and rationale of the network, the characteristics of its members, and the projections of future activities are described in earlier reports, and will not be repeated here. Rather, we concentrate on documenting the systems analysis and program planning that has been the bulk of the work, and on the resulting services now available under pilot operation.

Summarizing the report contents, Section 2 deals with our initial work in Task 2, catalog data file searching, and explores the problems and requirements of file directories. A seven part development plan for file searching is described, and methods for item selection and directory coding are advanced.

Section 3 concerns Task 3, Acquisitions Processing. The data obtained from a six-library survey of network members is documented and summarized. This data is serving as a principal design aid in arriving at a computer aided acquisitions system that will meet the needs and constraints of all network members.

Section 4 concerns the demonstration of services carried out under Task 4, or pilot operation. The production system is described problems are noted, and means for system improvement are given.

Much of technical detail of project activities is contained in the appendices: Appendix A contains the system flow charts of the acquisition systems presently operating in the six member libraries.

Appendix B contains the cost questionnaire which was sent to member libraries to determine the present costs of those products and services that the network will ultimately provide. Five of the

six questionnaires have been returned, and analysis of the data will begin shortly. The objective is to be able to make meaningful comparisons between manual and automated systems.

Appendix C describes an algorithm for dividing the L.C. Call Number. A program exists, but makes mistakes on certain unusual call numbers. The new algorithm is being subjected to manual testing prior to programming.

Appendix D contains an analysis of the MARC II format, which will become the new L.C. standard in July 1968, and which will require the writing of a conversion program for NELINET usage of MARC II.

Appendix E lists production problems, chiefly those of format, identified by the UNH staff during the shakedown of the pilot operation. A regular system of "problem reports" has now been initiated to identify and correct such problems promptly.

Appendix F is a manual of operating procedures for remote service requests. This has been developed as a guide for NELINET library personnel.

Appendix G describes comparative techniques for compression coding of entries, for use in file directories. A new technique, Alphacheck, is described which appears to offer the unusual combination of readability plus file discrimination.

Appendix H is a collection of sample NELINET products: catalog card sets, book labels, and book pocket labels.

2. CATALOG DATA FILE SEARCHING: PLAN FOR EXPERIMENTS AND SYSTEM DEVELOPMENT

2.1 INTRODUCTION

Our initial objective in the area of file searching is to develop methods for author and title access to cataloging records in large files. That is, search in support of cataloging that will supplement the access by LC Card Number that is now available. The cataloging records of interest are LC MARC records.

We plan to progress from an application where query data is quite accurate (search in support of cataloging; book generally in hand), to an application where the query data can be less certain (search in support of acquisitions), to later applications where query data is often decidedly imprecise (search in support of research). There will obviously be a large amount of carryover from one application to the next.

We describe here our plan for the development of a file searching system in support of cataloging, and present our initial approaches to file directory creation. We are specifically concerned in this report with the selection of items from the master cataloging record that will be directory elements, and the efficient coding of these elements for machine retrieval. New techniques are presented for the selection of "significant" title words for directory entry, and for the compression coding of author and title entries.

2.2 DEVELOPMENT PLAN FOR FILE SEARCHING

There are perhaps seven parts to the development of a file searching system, which we shall consider in the following order:

1. Selection of Directory Items: What items the directories should contain; that is, what items of the master catalog records should be represented outside the master file as an aid to master record location, and which items will be available only by accessing the master record itself.

2. Subset Selection: How item subsets can be selected from the master record when not all components of a tagged item are wanted, (e.g., selected title words).

3. Representation: How the selected items should be represented, i.e., coded or natural language, fixed or variable length.

4. Directory Design and Organization: How the directory elements should be strung together or associated in machine storage to minimize access time and storage requirements; and what percentage of the total file should be allocated to directories.

5. Query Language: How questions should be addressed to the file.

6. Search Strategy: What matching procedures should be applied to the input query and directory and file elements.

7. Machine Response: What man machine interactions are required, and what output responses should occur as a result of searching.

In this report we concentrate largely on items 1 through 3, in the context of title and author searching.

2.3 GENERAL SEARCHING METHOD

Figure 2-1 shows the general searching method, somewhat simplified from the standpoint of file structure. A search request contains an item (file access item) and verification data (other items used to verify that the selected record is the desired one). For example, a search request by personal author might use the title as verification, and vice versa. Either author requests or title requests might use the number of pages and the date of publication as verification items. We will also consider the inclusion of further verification data in the key itself, which may save time by eliminating needless record extraction and item matching.

The prime request parameter (e.g., author name) is transformed into a file key by computer. The same transformation process is used to generate the file keys in the file key directory, which are paired with the record address, and sorted by the key to form the directory.

At search time, the file key resulting from the request is matched against the file key directory, and any matches result in the corresponding catalog records being extracted for further processing. The further processing may include a matching of the uncoded request item and the uncoded record item, and similar comparisons of verification data in uncoded form.

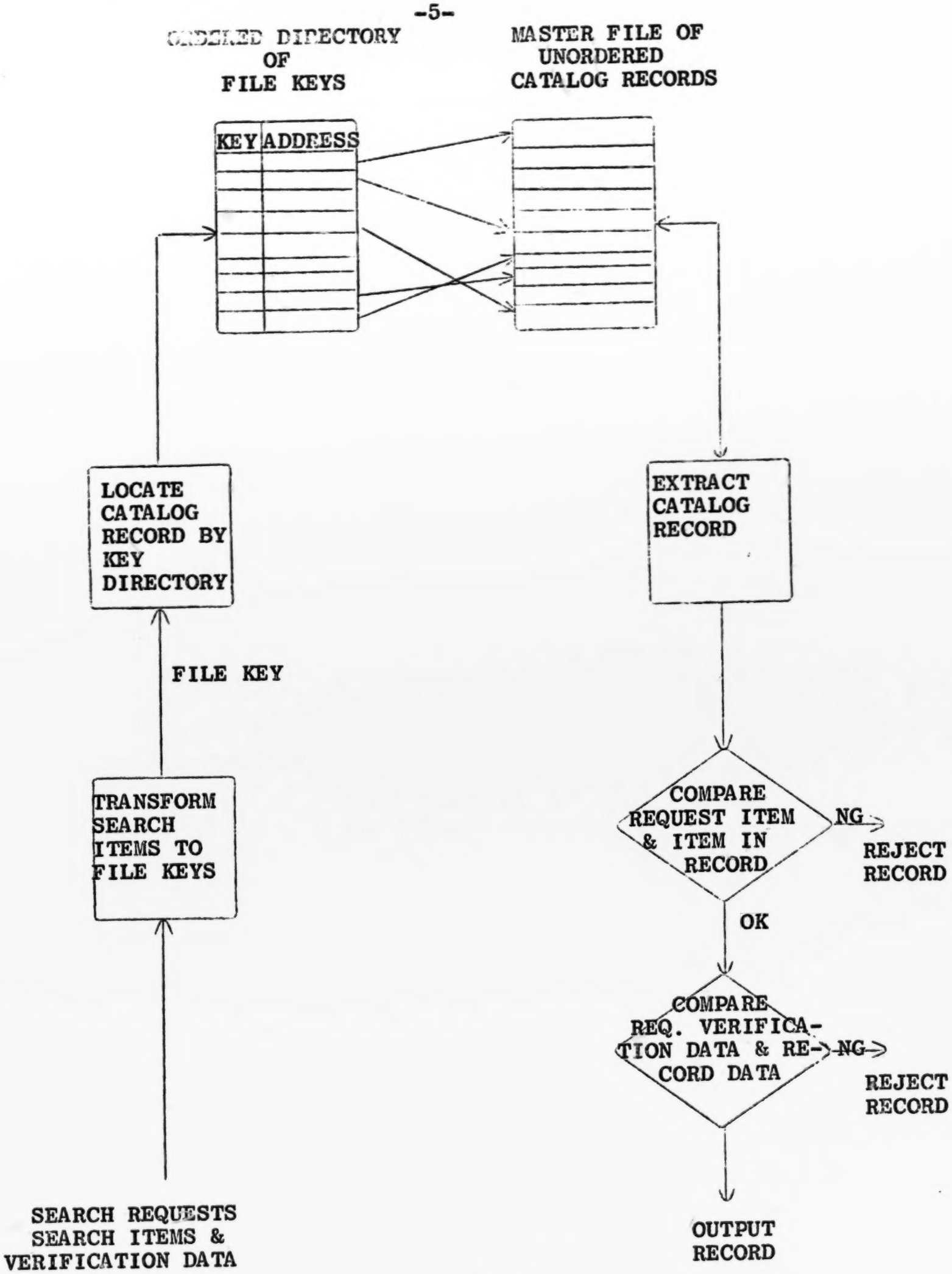


Figure 2 - 1

GENERAL SEARCHING METHOD

A prime objective of our initial retrieval tests is to establish an optimum form of file key for computer use. For discrete terms in the file that are chosen to be file access items (e.g. personal author, title), it is required to have an algorithmic method that may be used to operate on the item as it appears in the file, and transform it into an abbreviated form, or file key. These file keys would then be ordered to form a file directory, (or a hierarchy of directories). A directory is simply a list in file key order that contains the file key and the address of the master file record that the key item was extracted from (or the address of a higher order directory).

The function of the file key is to provide a simple surrogate of the item, generally short and fixed length, that is easy to sort and easy to match. Ideally, the method should:

1. Have sufficient resolution, or discriminating power, to result in different keys for different items.

2. Result in identical keys, or at least permit matching, from the predictable variations of the same entry.

3. Follow the general policy in matching, that missed records are intolerable, but a small number of incorrectly selected records are tolerable since these are easily rejected by criteria other than the primary key.

In brief, then, we are seeking efficient, short, fixed length compression codes for file access items. We begin by considering title and personal author searching in support of cataloging. We assume that the book is in hand, so that the form and spelling of query items are correct, and that difficult entries have been established where required. As we progress, we will be investigating retrieval methods that bypass the conventional establishing procedures and that can operate on imperfect query data.

2.4 TITLE SEARCHING

The elements of title searching consist of:

1. deriving a key from the title and other elements to be included in the title key, and placing this key in a directory;

2. determining a similar key from a search query;

3. matching these two keys by a suitable procedure.

2.4.1 Title Key Determination

We first consider the design of an algorithm for determining the title key, and start with that portion of the key

derived from the title. This algorithm will consist of two parts:

1. A process for selecting title words by eliminating words that do not contribute significantly to title discrimination and then reducing the remaining title words until a selected number remain. In this case, we will select a maximum of three words.

2. A process for converting the selected words of the title into a fixed length key.

2.4.11 Title Word Selection

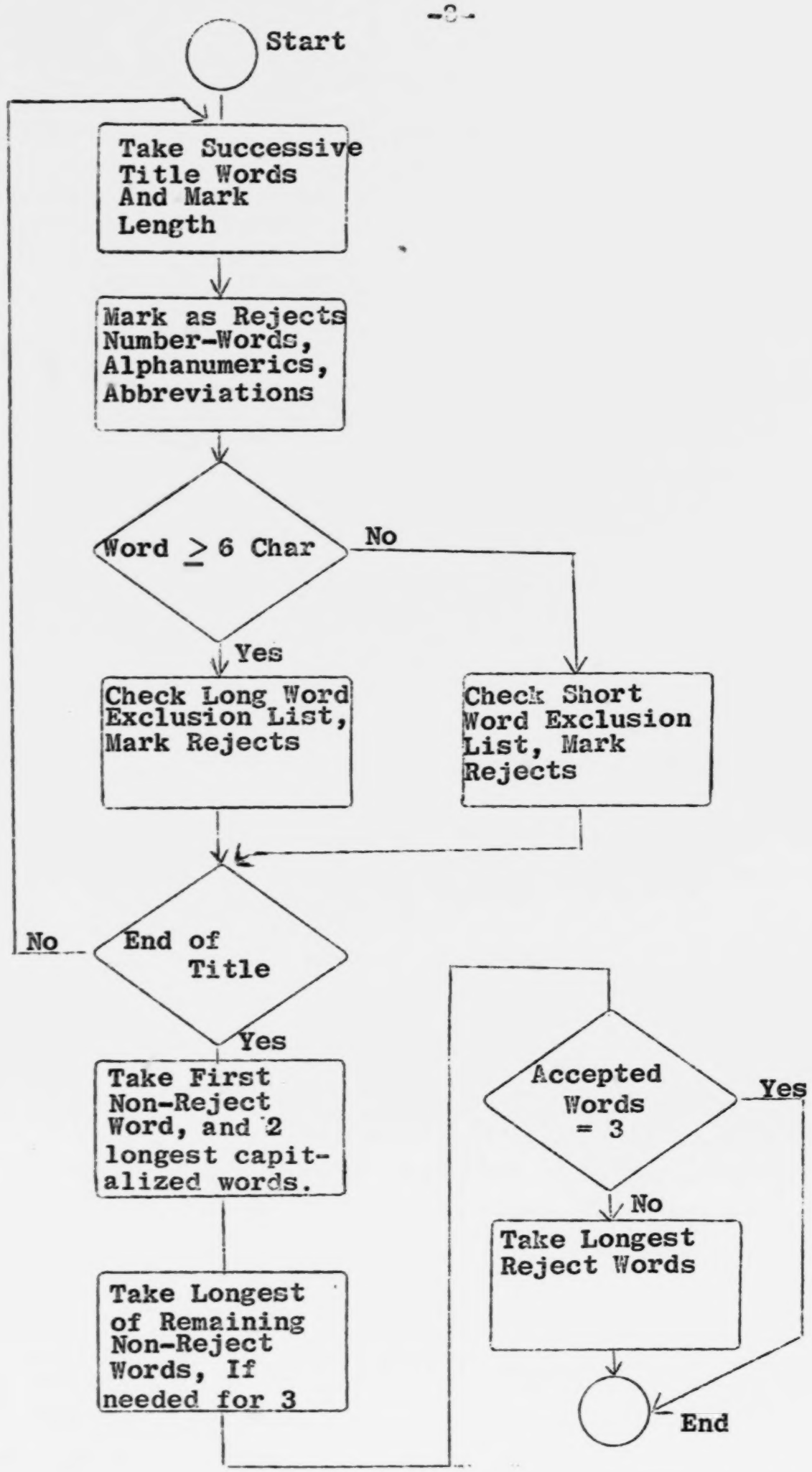
Part 1 is performed by the process shown in Figure 2-2. Successive words are marked as to their length, checked against one of two exclusion lists, and marked as a reject if they match an exclusion word. Three words are sought. The first non-reject word is taken; then longest non-initial proper names, if any; then the longest remaining non-rejects are taken. If the whole title is rejected, the three longest rejected words are accepted. Selected words are then placed in title order, forming the principal title key. The short word exclusion list contains the 99 most common English words as listed in the cryptography study of Gaines [1]. Of the hundred most common, only one has six letters ("should") and this is placed in the long word exclusion list. The long word exclusion list (Table 2-3) contains common non-discriminating title words (most frequently in or near initial position) that were selected by a study of title entries in Books in Print.

What we refer to as words are alphabetic elements only. Punctuation and capitalization are ignored (except for the detection of proper names in titles), and numeric elements are classed as rejects to minimize their use. The same applies to abbreviations. The occurrence of numerics and abbreviations as important factors in titles poses the conventional problems of file ordering and matching. The absence of "file as if" statements in the MARC records makes both abbreviations and numerics rather unmanageable as useful file keys at present. Hyphenated words are joined. Ties in word length are broken by a priority based on order of occurrence.

Table 2-1 illustrates the word selection process and key coding of titles selected from Books in Print. Initial articles have already been dropped in the sample source, but these would have been eliminated anyway, either by the short word exclusion list, or by a separate detection in the last program box of Figure 2-2.

Key Coding

The three selected title words are then converted into fixed length file keys by a suitable compression method.



TITLE WORD SELECTION PROCESS

FIGURE 2 - 2

TABLE 2-1

Title Word Selection and Key Coding

	TITLE	SELECTED KEY WORDS	TYPICAL FILE KEYS (6 character compression code)
1.	Three critical years; 1904-05-06	three critical years	THREE CIICAL YEARS
2.	3-5-7 minute talks on freemasonry	minute talks freemasonry	MINUTE TALKS FMSNRY
3.	Three little rabbits with queer little habits	three little rabbits	THREE LITTLE RBBITS
4.	Theory of papal monarchy in the 13th century	papal monarchy century	PAPAL MNRCHY CNTURY
5.	Theory of max-min and its application to weapons allocation problems	maxmin weapons allocation	MAXMIN WAPONS ALLCIN
6.	Introduction to Cobol computer programming for accounting and business analysis	cobol programming accounting	COBOL PGMMNG ACCUIG
7.	Introduction to the history of the science of politics	history science politics	HSTORY SCENCE PLTICS
8.	How we started out building our own home and went on to make a million in the construction business	started building construction	SARTED BLDING CSRUIN
9.	How to converse with spirit friends	converse spirit friends	CNVRSE SPIRIT FIENDS
10.	72 sure fire ways of having fun	sure fire having	SURE FIRE HAVING

TABLE 2-1 CONTINUED

	TITLE	SELECTED KEY WORDS	TYPICAL FILE KEYS (6 characters compression code)
11.	1984	1984	1984
12.	Physics	physics	PYSICS
13.	Introduction to Physics	introduction to physics	IRDUIN TO PYSICS
14.	Understanding put and call options	put call options	PUT CALL OPIONS
15.	Studien über hebraische Morphologie und Vokalismus auf Grundloge der mercatischen Fragmente der zwerten Kolumne der Hexapla des Origenes	studien morphologie vokalismus	SUDIEN MRPGLI VKLSMU
16.	Studies in folklore in honor of distinguished service professor Stith Thompson	folklore stith thompson	FLKORE STITH THMPON
17.	Ant and bee	ant and bee	ANT AND BEE
18.	U.S. versus Curtiss-Wright Export Corporation	versus curtisswright corporation	VERSUS CUSSWG CRPRIN
19.	Snail control in the prevention of biharziasis	snail prevention biharziasis	SNAIL PVNION BHRZSS
20.	As you like it	as you like	AS YOU LIKE
21.	Rajasi Janakananda	rajasi janakananda	RAJASI JNKNND
22.	Milton's theory of poetry and fine arts	miltons poetry fine	MLTONS POETRY FINE

TABLE 2-1 CONTINUED

TITLE	SELECTED KEY WORDS	TYPICAL FILE KEYS (6 character compression code)
23. Millimicrosecond pulse techniques	millimicrosecond pulse techniques	MLMCND PULSE TCIQES
24. Million random digits with 100,000 normal deviates	million random deviates	MLLION RANDOM DVATES

TABLE 2-2

Exclusion List of Shorter Words of Title
(99 Most Common English Words-After Gaines)

A
ABOUT
AFTER
ALL
AN
ANY
AND
ARE
AS
AT
BE
BEEN
BUT
BY
CAN
COULD
DO
EVEN
FIRST
FOR
FROM
GREAT
HAD
HAS
HAVE
HE
HER
HIM
HIS
I
IF
IN
INTO
IS
IT
ITS
LIKE
MADE
MAN
MAY
ME
MIGHT
MORE
MOST
MUCH
MUST
MY
NEW
NO
NOT
NOW
OF
OLD
ON
ONE
ONLY
OR
OTHER
OUR
OUT
OWN
SAID
SHALL
SHE
SO
SOME
SUCH
THAN
THAT
THE
THEIR
THEN
THEM
THERE
THESE
THEY
THIS
THOSE
TIME
TO
TWO
UP
UPON
US
VERY
WAS
WE
WELL
WERE
WHAT
WHEN
WHICH
WHO
WILL
WITH
WOULD
YET
YOU
YOUR

TABLE 2-3

Exclusion List of Longer Words of Title

Advance-	Essay	Making	Secret
Adventure-	Essential-	Manual	Selected
Anthology	Foundation-	Method-	Should
Application-	Fundamental-	Modern	Studies
Applied	General	Mystery	Technique-
Complete	Handbook	Nature	Textbook
Creative	Historical	Outline-	Theory
Critical	History	Practical	Treasury
Development	Introduction	Principle-	Treatise
Dictionary	Laboratory	Problem-	Understanding
Elements	Letters	Readings	

- ≡ Universal Symbol: Matches Null or any Symbol

The examples in Table 2-1 use the method of Recursive Decomposition, described in Appendix G. This method was devised by the author several years ago[2] and has been extensively tested. We discuss some newer methods, possibly superior for this application, in later sections and in Appendix G. Three adjacent six character codes in title order constitute the main title directory. The latter two words are used as individual headings for two permuted word directories containing the same three title words.

The choice of the coding system is conditioned by a desire to retain some readability in the keys while simultaneously achieving discrimination between similar words. Experiments in man-machine title searching will determine whether these keys can be useful as user feedback. It is possible that a longer character length may be desirable, and we will experiment as well with variable length word elements in a fixed length key space.

2.4.3 Title Directory-Preliminary

The title directory contains the three selected title word codes, an author code, the number of pages, and the address of the full cataloging record. The title directory is assumed to be the prime directory for search in support of cataloging. The inclusion of non-title items makes it possible to do secondary verification without accessing the master file. The balance of items to be included in directories versus those only to be available by master file access, is determined by a tradeoff between storage space and retrieval time. Figures 2-3 and 2-4 show examples of title directory entries. The author components, like the title components, are coded in a manner to retain some legibility for possible user feedback. The prime discrimination factor is obtained by the title words and their combinations; the remaining items serve as verification data, and occasionally as tie-breakers. The use of separate elements for the title code words, as opposed to a combinational one, permits partial matching or permuted query matching. The detailed matching procedures will be discussed in a subsequent memo; we are concerned here with the form and derivation of the directories.

2.5 AUTHOR SEARCHING

Searching by author is considered a secondary means of searching primarily used when title data is uncertain, or when title searches fail. We concentrate on personal author searching, but include corporate authors by the same means of word selection employed in titles.

Personal names are far less varied than titles, so that more powerful discrimination techniques are needed for their file keys. The basic problem, of course, is that for a given code word space, less than the natural language word space, how "good" can one make the code. If the objective is discrimination only, a randomized

Word 1 Code | Word 2 Code | Word 3 Code — Title Key

Surname Code | Forename Code | Initial — Author Key

Page Length

Address of Record in Master File

FNNGNS	WAKE □□	□□□□□□	JOYCE □	JAMES □	□	628	

Address

Directory Entry for: Joyce, James Finnegans Wake 628 p.

SNAIL □	PVNION	BHRZSS	WORLD □	ORGIZN	3	234	

Address

Directory Entry for: World Health Organization Snail control in the prevention of biharziasis 234 p.

(corporate entry subject to 2-word extraction by title rules, entered in order with number of words as last element)

GTNBRG	GALAXY	□□□□□□	MLUHAN	HRBERT	M	294	

Address

Directory Entry for: McLuhan, Marshall The Gutenberg galaxy 294 p.

(First establish entry: Marshall McLuhan ⇒ Herbert Marshall McLuhan)

IRDUIN	PBBLTY	APPINS	FELLER	WILLIAM □		461	

Address

Directory Entry for: Feller, William An introduction to probability theory and its applications 461 p. (note "probability" only non-reject word)

FIGURE 2-3

Title Key			Author Key		pages	Address
FINNED	WAKE□□	□□□□□□	JOYCE □	JAMES□	□ 628	
						Address

Joyce, James Finnegans wake 628 p.

SNAIL□	PREVEL	BIHARV	WORLD □	ORGANX	3 234	
						Address

World Health Organization Snail control in the prevention of biharziasis 234 p.

GUTENR	GALAXY	□□□□□□	MCLUHZ	HERBEN	M 294	
						Address

McLuhan, Herbert Marshall The Gutenberg galaxy 294 p.

INTROR	PROBAH	APPLIV	FELLER	WILLIP	□ 461	
						Address

Feller, William An introduction to probability theory and its applications 461 p.

FIGURE 2-4

Examples of Entries in Title Directory

(Compression by Truncation Plus Alphacheck)

binary code will be most efficient. If printout is occasionally desired, a randomized alphabetic code will do nearly as well. For viewing purposes, a selected letter code is best. A selected letter code with a check character can provide some of the best features of selected letter and randomized codes. We show six varieties of codes suitable for coding the author component of directories in Figure 2-5.

Code 1 is the method of Recursive Decomposition that we have used for title coding in the example of Figure 2-3. A six character code is selected. Code two is similar, except that the word has been decomposed to five characters, and an alphanumeric check character, generated by the Alphacheck method, has been appended at the end. For a given code space, the check character adds more discrimination than is lost by the removal of an additional character. In using Alphacheck, words of six characters or less are self-representing; longer words are compressed to five characters by a suitable selection algorithm; and the rejected letters are operated on to form the final alphanumeric check character.

Code three is a four character randomized code of extremely high discrimination, generated via Transition Distance Coding. Different character lengths can be chosen as required. By its nature, it has no obvious relation to the clear text form of the word.

Code four is the same code with the initial of the clear-text appended. The initial aids in file maintenance and in matching on the initials of forenames, etc. Code five is the phonetic Soundex code that unifies the codes of similar sounding names.

Code six is Alphacheck, using simple truncation with a final check character. This method shows great promise in combining readability and discrimination, and will be the preferred code if discrimination tests show it to be competitive with less readable codes.

These methods will be subject to evaluation and comparison during subsequent development of file searching methods. Our initial tests will evaluate these codes as applied to the title directory and the author directory. Alphacheck will be given special attention, since the computational requirements are moderate and the legibility is best. If this choice shows sufficient discrimination, then a more complex code may not be required. This will be determined.

2.5.1 Transition Distance Coding and Alphacheck

The Recursive Decomposition code was devised several years ago, and its discrimination capability is known. Transition Distance Coding and Alphacheck have been devised for this project and are as yet untested. Transition Distance Coding is a new technique for the generation of high discrimination fixed length alphabetic codes

	Surname Natural Language:	JOHNSEN	JOHNSON	JOHNSTON
1.	Selected Letter Compression Code: (Recursive Decomposition)	JHNSEN	JHNSON	JHSTON
2.	Selected Letter Compression Code with Alphanumeric Check Character: (Recursive Decomposition and Alphacheck)	JHSENL	JHSONL	JHSON5
3.	Pure Quasi-Random Code: (Transition Distance Coding)	BFTZ□□	DNWU□□	ZIKY□□
4.	Initial Ordered Quasi-Random Code: (Cleartext initial added to Transition Distance Coding)	JBFTZ□□	JDNWU□□	JZIKY□
5.	Phonetic Code: (Soundex)	J52	J52	J52
6.	Truncation Plus Check Character: (Alphacheck)	JOHNSV	JOHNSX	JOHNSD

FIGURE 2-5

Examples of Forms of Fixed Length Coding
Suitable for Directories

from variable length words. It may be classed as a randomizing or 'hash' code, but unlike many quasi-random codes it is sensitive to the letter sequence in words as well as the letter content. The method appears capable of a high randomization (and hence better discrimination) than many earlier techniques that have been quantatively tested.

Alphacheck is a reduced form of Transition Distance Coding that generates a single character from an arbitrary length letter sequence. When applied to the excluded letters of a selected letter compression code, it generates a single character surrogate of the missing components; and this appended to the compression code raises the discrimination of the code considerably.

Appendix G described both methods in detail.

2.5.2 Author Directory-Preliminary

Examples of entries in the author directory are shown in Figures 2-6 and 2-7. The entries contain the same items as the title directory, except that the author codes are in initial position. In Figure 2-6, the surname and forename have been coded into six characters by the method of Recursive Decomposition plus Alphacheck. If a name is six symbols or less, it is taken verbatim; if it is longer, a five symbol code is created by Recursive Decomposition and a one symbol Alphacheck code is generated from the omitted characters. In this process it will be observed that at least two characters will have been omitted wherever compression is necessary In Figure 2-7, all elements are coded by truncation plus Alphacheck.

Non-personal author names are operated on by title rules to extract the two "most informative" words, and these are placed in title sequence in the author position, along with the number of words in the corporate name which is placed in the middle initial position. This middle initial position, therefore, serves as a personal author - non-personal author discriminant depending on whether it is non-numeric or numeric, respectively. The subject of non-personal author retrieval will require separate study and separate approaches. The method here is merely one initial way to handle non-personal authors within a directory framework designed for personal authors.

2.6 CONCLUSION

We have considered here and in Appendix G several approaches to directory design, concentrating on item selection and coding. These approaches will be tested as we progress to the additional tasks of file organization, query language, and the other remaining aspects of catalog data file searching described in section 2.

Surname Code	Forename Code	Initial	Word 1 Code	Word 2 Code	Word 3 Code	Page Length	Address of Record in Master File

Author Key — Title Key

(Recursive Decomposition Plus Alphacheck) (Recursive Decomposition)

JOYCE □	JAMES □	□	FNNGNS	WAKE□□	□□□□□□	628	Address

James Joyce Finnegans wake 628 p.

WORLD □	ORGIZN	3	SNAIL □	PVNION	BHRZSS	234	Address

World Health Organization Snail control in the prevention of biharziasis 234 p.

MLUHAN	HRBERT	M	GTNBRG	GALAXY	□□□□□□	294	Address

Herbert Marshall McLuhan The Gutenberg galaxy 294 p.

FELLER	WILLIAM	□	IRDUIN	PBBLTY	APPINS	461	Address

William Feller An introduction to probability theory and its applications 461 p.

FIGURE 2-6

Examples of Entries in Author Directory

(Compression by Recursive Decomposition)

Author Key			Title Key			Page Length	Address
JOYCE□	JAMES□	□	FINNED	WAKE□□	□□□□□□	628	
							Address

Joyce, James Finnegans wake 628 p.

WORLD□	ORGANX	3	SNAIL□	PREVEL□	BIHARV	234	
							Address

World Health Organization Snail control in the prevention of biharziasis 234 p.

MCLUHZ	HERBEN	M	GUTENR	GALAXY	□□□□□□	294	
							Address

McLuhan, Herbert Marshall The Gutenberg galaxy 294 p.

FELLER	WILLIP	□	INTROR	PROBAH	APPLIX	241	
							Address

Feller, William An introduction to probability theory and its applications 461 p.

FIGURE 2-7

Examples in Entries in Author Directory
(Compression by Truncation plus Alphacheck)

REFERENCES

1. Gaines, Helen Fouche *Cryptanalysis* Dover, New York, 1965.

2. Nugent, William R. and Vegh, Alexander *Automatic Word Coding for Computer Language Processing* Report RADC-TDR-62-13 February 1962, Vol I: AD-272-401, Vol. II: AD-272-402.

3. Pratt, Fletcher *Secret and Urgent, the Story of Codes and Ciphers* Blue Ribbon Books, Garden City, New York 1939.

3. TASK 3. ACQUISITION PROCESSING

3.1 ANALYSIS OF EXISTING ACQUISITIONS SYSTEMS

During this report period the six participating libraries were visited and the acquisition procedures were studied in detail. The book selection activities and the ordering-accounting operations performed in the administrative offices of the university were not examined. The study concentrated on the processing of books or one-time orders. Information regarding the treatment of standing orders or serials was picked up only incidentally. As evidenced by the flow charts in Appendix A, the degree of detail gathered was not uniform in all libraries.

To facilitate comparison, significant points of variation are listed in Table 3-1, and the practice in each library noted. Although there may seem to be more variations than similarities, the variations are due to differences in method rather than differences in function. In each library, the goals are the same — to acquire materials for the library and to provide control over this operation. Both the library and the administrative offices of the university are involved in and exert certain controls over the operation. Since the six universities are state institutions, the states also exercise some control.

In each of the six universities, the library is responsible for (1) determining what should be ordered; (2) determining from whom it should be ordered; (3) preparing the orders; and (4) approving invoices for payment. All of the libraries keep records of expenditures for budget control purposes even though some of them receive statements from their university accounting offices.

In three institutions, the universities' administrative offices become aware of (or involved in) the acquisitions process at order time. In the other three institutions, the universities are not involved until the book and invoice have been received. Three of the universities are responsible for the writing of checks. The other three forward invoices to their state treasurer's office for payment.

3.1.1 Forms

All six libraries use a request form. The data elements preprinted on the forms are shown in Table 3-2. Five of the six libraries use a multiple copy library ofder form. Table 3-3 lists the data elements preprinted on these forms. All six libraries prepare a purchase order or invoice voucher at some point in the ordering-receiving operation. In three institutions these vouchers are university forms; in the other three institutions, they are state forms.

TABLE 3-1

Variations in Library Acquisitions Practices

	Library A	Library B	Library C	Library D	Library E	Library F
Multiple library order forms used	o	x	x	x	x	x
No. of parts in request form (± additions or deletions)	1(+1)	1	2(-1)	2(-1)	1	1
No. of parts in library order form	-	5	7	7	5	5
No. of parts in voucher (± additions or deletions)	4	6(+2)	5(-1)	5(-3)	5	1
No. of library order forms sent to vendor.	-	o	2	2	1	1
No. of vouchers sent to vendor	2	1	o	1	5(a)	o
Vouchers typed when item ordered	x	x	o	x	o	o
Vouchers typed when payment approved	o	o	x	o	x	x
Vouchers contain bibliographic data	x	x	o	o	o	o
Voucher forms are state forms	o	x	x	o	x	o
Voucher signed by university personnel when item ordered	x	x	-	x	o	-
Voucher signed by library personnel	x	o	x	o	x	x
Orders sent to vendor via Univ. Adm. Off.	x	o	o	x	o	o
Record of Order in Public Catalog	o	x	x	x	x	x
Public Catalog record filed when item ordered	-	o	x	x	x	x
Public Catalog record filed when item received	-	x	o	o	o	o
Records kept of encumbered funds	o	o	o	x	x	x
Library expenditure statements prepared by Univ. Adm.	x	x	o	o	x	x
Books held until invoices received	x	o	x	x	o	o
Requestor notified when item ordered	o	x	x	o	o	o
Requestor notified when item received	o	x	o	o	o	o
Requestor notified when item available	x	x	x	x	x	x
Standing orders for Series	x	x	x	x	x	x
Standing orders to Publishers	x	x	x	x	x	o
Standing orders to Vendors	o	o	x	x	x	o
Processing done for outside libraries	o	o	o	x	o	x
Accession list prepared	x	x	o	x	o	o

x = Yes
o = No
- = Doesn't apply
(a) = These are blank forms for vendor to fill in

TABLE 3-2

Request Form — Preprinted Data

	Library A	Library B	Library C	Library D	Library E	Library F
Author	x	x	x	x	x	x
Title	x	x	x	x	x	x
Edition			x	x	x	x
Place			x	x	x	x
Publisher	x	x	x	x	x	x
Publisher's Address		x				
Date	x	x	x	x	x	x
Series			x	x	x	x
Volumes			x	x	x	x
Destination			x		x	
Reserve	x					
Course no.	x					
No. of students	x					
Requestor	x		x	x	x	x
Department	x	x			x	x
No. of Copies			x	x	x	x
Price	x	x	x	x	x	x
Vendor	x		x	x	x	x
Catalog no.				x	x	x
Item no.				x	x	x
Fund			x	x	x	x
Approved by			x	x		
Notes			x	x	x	

TABLE 3-2 CONTINUED

	Library A	Library B	Library C	Library D	Library E	Library F
Hold	x					
Searched in					x	x
Searched in year					x	
Searched in vol(s)					x	
Public catalog	x		x	x	x	
Order file	x		x		x	
Serials file	x				x	
CMC	x					
ABC			x			
Verified in					x	
Verified in year					x	
Verified in vol(s)					x	
Verified in page					x	
PTLA				x		
BBIP				x		
BIP			x	x		
CBI			x	x		
BPR			x	x		
BNB	x					
LJ	x					
PW	x					
KV				x		
LV				x		
PBIP				x		

TABLE 3-2 CONTINUED

	Library A	Library B	Library C	Library D	Library E	Library F
UM				x		
LC				x		
FB				x		
ULS				x		
NST				x		
Series checked in Cat.				x		
SSL				x		
SO				x		
ULS				x		
NST				x		
Added copy				x	x	x
Added volume					x	x
Added edition					x	x
Continuation					x	x
Serial					x	x
Order no.	x		x	x	x	x
LC card no.	x		x	x	x	
Call no.			x	x	x	x
Date requested				x	x	x
Date ordered				x	x	
Date received				x	x	
Rush						x
Sold/out of print					x	

TABLE 3-2 CONTINUED

	Library A	Library B	Library C	Library D	Library E	Library F
Bibliographer				x	x	
Cited in				x		
Cost				x		
Date of invoice				x		
Special instructions				x		
Quotation	x					

TABLE 3-3

Library Order Forms — Preprinted Data

	Library B	Library C	Library D	Library E	Library F
Author	x	x		x	
Title	x	x		x	
Place	x	x	x	x	x
Publisher	x	x	x	x	x
Year	x	x	x	x	x
Edition	x	x	x	x	x
Series	x		x	x	
Volumes	x	x	x	x	x
Requestor	x	x		x	x
Fund	x	x	x	x	x
Cost	x			x	x
LC card no.	x	x	x	x	x
Classification no.	x			x	
List price	x	x	x	x	x
Date ordered	x		x	x	x
Order no.	x	x	x	x	x
Dealer	x	x	x	x	x
No. of copies	x	x		x	x
Date received	x			x	x
Department					x
Catalog to		x			
Item			x		

3.1.2 Files

As Table 3-4 illustrates, there are many variations in the files kept in the six libraries to control the acquisitions process. These files differ in:

(1) The data contained in each record. (Indicated in Table 3-4 by R (Request Form), L (Library Order Form), V (University or State Purchase Order or Invoice Form) and I (Dealer's Invoice).

(2) The arrangement of the file.

(3) The time at which the record is filed.

(4) The time at which the record is pulled.

The first entry in Table 3-4 should be interpreted as:

> Library A has one file arranged by main entry. The data contained in this file is the data recorded on a request form plus Library of Congress cataloging copy. It is filed when the item is ordered and removed when the item is is ready for circulation. The file is located in the Acquisitions Department.

The desirability of multiple access points to the acquisitions record is obvious from the variety of file arrangements in the six libraries. Five of the six libraries also file a record of the order in the public catalog thereby providing another file arranged by main entry. This remains in the catalog until the item has been completely processed and cards are filed.

Of particular interest to the design of a machine based system is the arrangement of files by title in two of the libraries. The Oak Park, Illinois processing center also reports that they have changed the arrangement of their order file from author to title because their "clerical staff finds it easier to work with."[1] Searching by title may also prove to be better suited to a computer than searching by main entry.

Also apparent in Table 3-4 is the number of files which must be checked to determine the status and location of an item in process.

3.1.3 Tie-in with the Cataloging Process

In order to integrate the Center's acquisitions and cataloging services, the routines in the six Cataloging Departments were

---

1 Hendricks, Donald. Organization for Processing at the Book Processing Center, Oak Park, Illinois. Library Resources and Technical Services, vol. 10, Fall 1966, pp. 479-489.

TABLE 3-4

Comparison of Acquisitions Files

(File is located in Acquisitions Department unless noted otherwise.)

File In-Out Times

1. Item ordered
2. Item received
3. Library approves payment
4. University approves payment
5. LC cards ordered
6. LC cards received
7. Item available
8. End-of-month

Form Filed

R Request form
L Library order form
LC LC cards
LCP LC proof
V Voucher
I Invoice

Arrangement	Library A		Library B		Library C		Library D		Library E		Library F	
	In-Out	Form	In-Out	Form	In-Out	Form	In-Out	Form	In-Out	Form	In-Out	Form
Main entry	1-7	R[a]	1-7	L			1-3	R+L[a]	1-2	L[b]	1-2 3+	R+L[a] L[b]
Title					3-6 1-3	L[b,c] LCP[b]			1-2	R		
Vendor	1-3	V	1-2	V	1-2	R+L			1-2 2-3	L L	1-2 2-3 3-4	L L I
Fund							1-3 [illegible]	L L				
Department							1-3	L				
Yr(s), main ent.	7+	R	7+	L[b]			3+	R			3+	L
Yr(s), vendor	3+	V	3+	V[c]					3+ 4+	L V	4+	I
Date, main ent.									5-6	L[b]		
No., lib. ord.							1-3 3+	L L				
No., voucher			1-3 4+	V[c] V[c]			1+	V[e]				
No., LC card			5-6 6+	L[b,f] L[b,g]								

a. Also contains LC copy
b. In Catalog Department
c. In Library Adm. Off.
d. Or "6-3 LC"
e. Voucher no. is same as lib. ord. no.
f. If no card no., arranged by date, main entry
g. If no card no., arranged by main entry

reviewed with special attention given to the procurement of Library of Congress cataloging copy. The findings are given in Table 3-5.

3.2 DESIGN OF THE MECHANIZED ACQUISITIONS SYSTEM

The requirements for the centralized acquisitions system are being formulated; a master list of data elements is being prepared; and various file organizations are being considered.

TABLE 3-5

Acquisition of Library of Congress Cataloging Copy

	A	B	C	D	E	F
Subscribe to LC proof slips or Title II cards	x	o	x	x	x	o
File proof slips by main entry	x	-	o	x	x	-
File proof slips by title	o	-	x	o	o	-
Order LC cards	x	x	x	x	x	o
Order LC cards at order time	x	o	x	o	o	-
Order LC cards at receipt time	x	x	x	x	x	-

4. TASK 4: DEMONSTRATION OF SERVICES

Demonstration of services began in December 1967. Since then, the University of New Hampshire has been sending requests for cards and labels via dataphone to the Cambridge office of Inforonics. These requests have included the LC card number and the local copy, volume, and branch information needed to produce cards and labels. A number of problems were encountered.

The proportion of requests found on the MARC tapes was disappointing. Initially requests were matched with only a single MARC tape. Three more tapes have been acquired and converted to master file format. They should bring the proportion of requests found somewhere near the 50% which the University of Toronto experiences (Computers and the humanities 2:104, Jan. 1968).

Appendix E lists a number of problems, chiefly format ones. and the decisions about them based on a meeting in Durham on February 8.

The patience with which New Hampshire has endured the headaches of being first deserve acknowledgement. They have requested much and received little. This helped shake-down the system but was not very fruitful for New Hampshire. The situation is improved now; useful products are being received.

Delivery by bus package express as an alternative to parcel post has been investigated. It seems feasible for several libraries:

BOSTON to:	BURLINGTON	DURHAM	AMHERST	STORRS
8:30 am	2:30 pm			
9:00 am		11:05 am		
10:30 am		12:20 pm		
9:15 am			12:15 pm	
10:30 am			2:25 pm	
1:15 pm			4:15 pm	
9:15 am	(via Springfield, Mass.)			1:05 pm
1:15 pm	(via Springfield, Mass.)			4:35 pm

In order to gain data on the libraries' present processing costs for comparison with those of the NELINET operation, a questionnaire was designed. (See Appendix B) Completed questionnaires are being received and data from them will be included in a future report.

A NELINET Newsletter has been started for brief, informal communications among the participating libraries.

In December, a meeting was held at the University of New Hampshire to introduce the request form. At it the pamphlet Operating procedures for information transmittal in the New England Library Information Network was used to familiarize their catalogers and clerks with the procedures for requesting cataloging materials. (See Appendix F for the pamphlet.) The low frequency of errors in the requests reflects well upon all concerned.

During the period July 1967 to Jan. 1968, the programs written for Task I (data file creation and catalog card production) have been constantly revised. These revisions were needed to correct misinterpretations of MARC I documentation, to handle requests from several libraries more efficiently and to satisfy the libraries' needs when they differed from the Library of Congress format. These problems are fairly well overcome now and production using requests from more than one library is about to begin. Rhode Island will be the second library.

MARC I data tapes were supplied by Widener Library Computer Center, Harvard University. As the MARC information was scattered on four tapes and each tape overlapped serially, the Widener Library sorted, merged and reduced the tapes from 556 bits per inch to 200 BPI. This gave us a data base of 28,500 bibliographic entries with which to test the system. Initial trials had been run with a data base of 8,000 MARC records, also supplied by the Widener Library.

Figure 4-1 shows the present configuration of the system for request and production of cataloging materials. As soon as certain improvements have been incorporated, the system configuration will be that shown in figure 4-2. These improvements will increase efficiency in three ways:

First, search time will be reduced by merging several libraries' requests, matching them with the MARC tapes and then sorting them back into order by library. As the number of libraries and the number of MARC tapes grow, this will become extremely important.

Second, use of the 120 character line printer at the Harvard Computing Center will greatly reduce the time required for card production while improving the quality of the printing.

Third, making a list of requests not found on the MARC tapes in teletype code will facilitate reporting to the libraries. This should reduce if not prevent formation of a "NELINET backlog" at the local libraries.

While the system will be a "batch" processing rather than an "on-line" one even after these improvements, it is expected that

completed products will be faster than current local processing.

Sample outputs are enclosed as Appendix H. These are representative in that they show single and multiple card output, with appropriate subject and added entry tracings overprinted as headings.

Beginning in July 1968, the MARC record will appear in a totally new format. (See Appendix D for a comparison of MARC I and MARC II.) For a realistic demonstration of NELINET services to continue after that date, either new search and print programs must be written or a means of making the MARC II tapes compatible with the present programs must be found. While the former will probably need to be done eventually, the latter seems a wise short-run expedient. One thing certain about MARC II is that it will change; "collapsing" the tapes into the NELINET version of MARC I while MARC II matures should minimize our reprogramming.

To allow the libraries access to the national teletype network, the dataphones used to communicate with Inforonics were provided with teletype capability. (Dataphone lines are preferred for use with computers.) This was done at minimal additional cost. One librarian estimated that teletyping halved the time needed to find a title they wanted to borrow on interlibrary loan. A major reason was that they could quickly ask a second library for a title when they learned so much sooner the first did not have it.

In summary, cooperative mechanization in the New England state university libraries is rapidly becoming a reality. The New England Library Information Network is presently producing catalog cards and labels with local data (branch, oversize, etc.) on them from the MARC tapes for one member. It will soon do so for all the others.

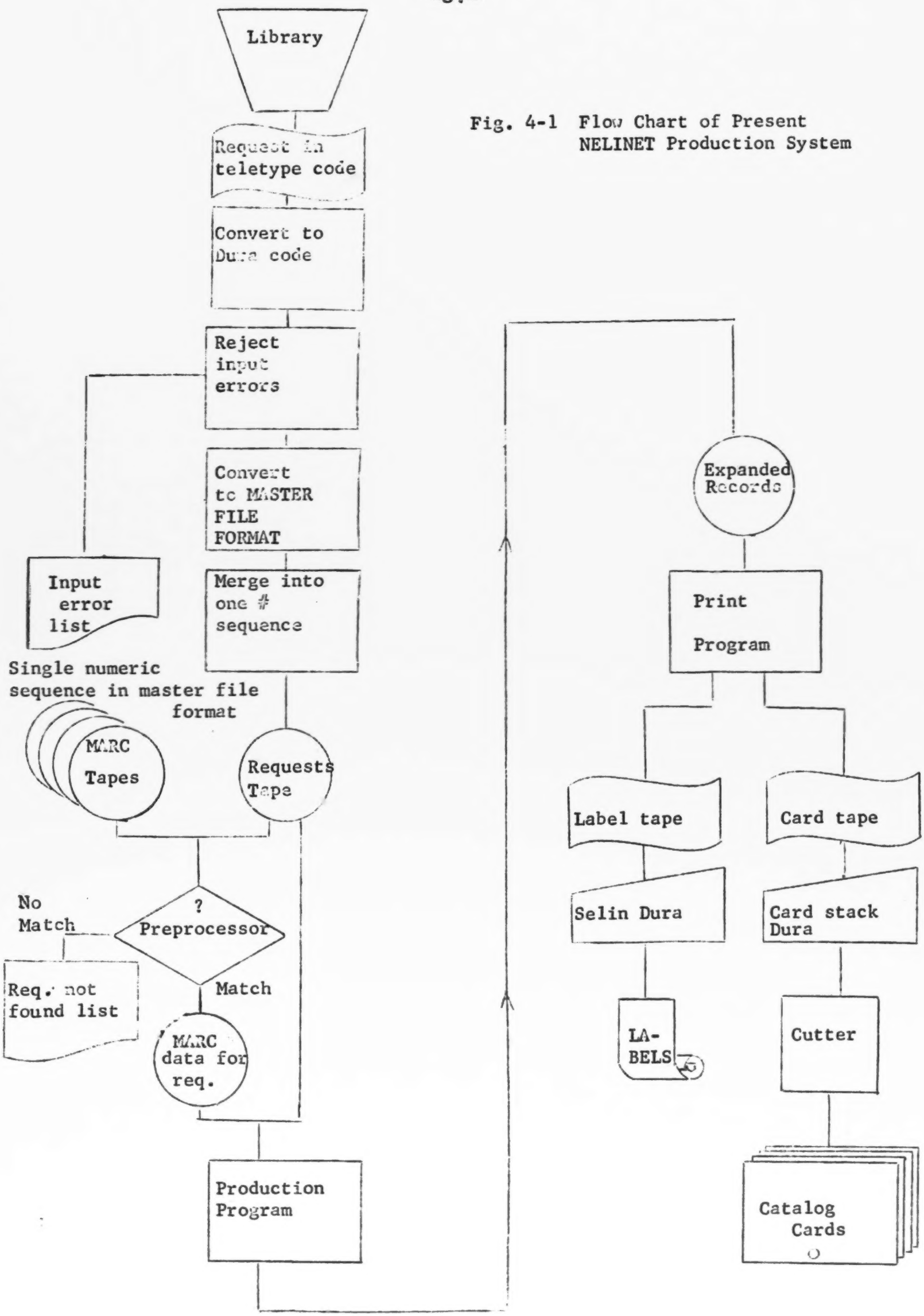


Fig. 4-1 Flow Chart of Present NELINET Production System

Fig. 4-2 Flow Chart of Improved NELINET Production System